

# A marked renewal process model for the size of a honey bee colony

Aihua Xia<sup>a,1,\*</sup>, Richard M. Huggins<sup>a</sup>, Martine J. Barons<sup>b,2</sup>, Louis Guillot<sup>a</sup>

<sup>a</sup>*School of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria, 3010 Australia*

<sup>b</sup>*Department of Statistics, The University of Warwick, Coventry CV4 7AL United Kingdom*

---

## Abstract

Many areas of agriculture rely on honey bees to provide pollination services and any decline in honey bee numbers can impact on global food security. In order to understand the dynamics of honey bee colonies we present a discrete time marked renewal process model for the size of a colony. We demonstrate that under mild conditions this attains a stationary distribution that depends on the distribution of the numbers of eggs per batch, the probability an egg hatches and the distributions of the times between batches and bee lifetime. This allows an analytic examination of the effect of changing these quantities. We then extend this model to cyclic annual effects where for example the numbers of eggs per batch and the probability an egg hatches may vary over the year.

**Keywords:** Bee colony size; Renewal process; Stationary distribution.

---



---

\*Corresponding author

*Email address:* aihuaxia@unimelb.edu.au (Aihua Xia)

<sup>1</sup>Research supported by ARC Discovery DP150101459.

<sup>2</sup>Research supported by EPSRC grant EP/K039628/1.

## 1. Introduction

The world population reached 7.3 billion as of mid-2015, indicating that the world has added approximately one billion people in the last twelve years and is projected to pass 9 billion by 2050 (UN, 2015). It is vital, therefore, that optimal use is made of the world's finite resources for food production (Collier, 2009). Whilst honey bees are not the only pollinators (Rader et al., 2016), it is estimated that 70% of important food crops are bee-pollinated (Datta et al., 2013) and so the status of pollinators, and bees in particular, is of a key concern in global food security (Lonsdorf et al., 2009; Blaauw and Isaacs, 2014). Many large-scale agricultural businesses now employ peripatetic bee services in order to ensure adequate pollination of crops (Gordon et al., 2014; Bishop et al., 2016). There is evidence that pollinator populations are declining and this is a cause for concern for policy-makers (Hafi et al., 2012; Breeze et al., 2012; Vanbergen et al., 2014; Chauzat et al., 2014; Potts et al., 2009). This observation motivates our development of a probabilistic model for the colony size. The model allows us to examine how the size depends on the distribution of the numbers of eggs per batch, the chance of hatching and the distributions of the times between batches and bee lifetime. In turn this gives insight into how a disease for example that affects one or more of these quantities affects the colony size.

The dynamics of a bee colony are complex. After her mating flight, the queen remains in the hive laying eggs, fed and cared for by worker bees. The number of eggs laid are influenced by nectar flow into the colony, which in turn, is influenced by the number of bees out foraging and seasonal effects on availability. During winter, in the absence of forage, there is little or no egg-laying and the colony relies on the honey and pollen stores accumulated during the summer. In this situation the lifetime of a bee can be significantly longer so that the colony can survive winter (Mattila et al., 2001). Eggs hatch into larvae which pupate after a few days of feeding (Collins, 2004) and emerge as juvenile adult bees about three weeks after the egg is laid and living a further 30-50

days. Juvenile adult worker bees mature as they work within the hive until they become foragers after about three weeks after emergence. Colony sizes can range from 20,000 to 100,000 (Goodman and Kaczynski, 2015; Owen, 2015). Pressures on a colony include short and long term weather effects, for example, the impact of the season on the hatching rate through changes in relative humidity was observed in (Al-Ghamdi et al., 2014). Other effects include diseases (Arundel, 2011; Gordon et al., 2014; Bull et al., 2012; Fürst et al., 2014), pests (Chandler et al., 2000; Cook et al., 2007; Ryabov et al., 2014), predators (Capri et al., 2013), pesticide, fungicide and herbicide use (Botias et al., 2015; Dively et al., 2015; Godfray et al., 2014; Pettis et al., 2013). When these pressures reduce the number of mature foragers, juvenile adults become precocious foragers, that are less efficient and shorter-lived than mature foragers (Schulz et al., 1998). This can lead to collapse of the colony (Perry et al., 2015). We believe that a necessary first step in understanding the dynamics of bee colonies is to understand their behaviour in a stable environment.

The literature contains both probabilistic models examined by simulations and deterministic mathematical models that can be used to examine the effects of environmental changes and disease on colony dynamics (Becher et al., 2013; Khoury et al., 2013; Lever et al., 2014). These models are themselves complex and, whilst the effects are evident in the results, they often give little understanding of the mechanism. For example, the model description for the simulation program BEEHAVE is comprehensive (Becher et al., 2014) but it is complex and mathematically intractable. Our contribution is a parsimonious and analytically tractable probabilistic model for the daily colony size which captures the key features and behaviour of the natural system under stable conditions. This is the first step in the development of more complex models but is of interest as we demonstrate the existence of a stationary distribution for the colony size. In particular, the mean of the stationary distribution is simply expressed in terms of the mean of the batch size, the probability that eggs hatch into bees and the means of bee lifetime and time between batches. Our model

can also be viewed as a type of batch immigration-death model: bees arrive in batches of eggs laid by the queen and subsequently die.

The birth and death process with immigration is well-known in the random process literature (Grimmett and Stirzaker, 2001, pp. 276–278; Lawler, 2006, p. 76). Classical immigration-death process models are first order Markovian. That is, given the population size at time  $t$  the population size at time  $t + 1$  is independent of the past of the process. In our case the daily death rate of the bees depends on the age structure of the colony and is not first order Markovian in this sense. In theory if there is an upper bound on bee lifetimes one could obtain a Markov property by extending the state space but we consider a more direct approach based on marked renewal processes.

Our approach requires modelling the distributions of four quantities. The first is the time in days between the days on which the queen lays batches of eggs. Often this will be one day but our results allow this to have a distribution over the positive integers, with daily layings being a special case. The second is the number of eggs in a batch. The third is the probability an egg successfully hatches into a bee. The fourth is the lifetime of a bee. This initial model represents a perfect bee world without seasonal effects where there are always sufficient resources so that the batch sizes on different days have the same distribution and the probability that an egg hatches into a bee remains the same. We then extend this to a cyclic model where the mean batch size and the probability of an egg hatching vary over an annual cycle which allows us to consider seasonal effects.

Our initial hypothesis was that for the given distributions of the time between batches, the number of eggs in a batch, the indicator whether an egg successfully hatches and the lifetime of a bee, with these quantities being mutually independent, there is a stationary distribution for the size of the colony.

The colony size is modelled as a marked renewal process with the number of batches being a renewal process and the number of bees hatching in each batch being the mark. Then coupling techniques (Lindvall, 1992, pp. 21-34) can be used to study its stationary distribution. In Section 2 we describe the model and in Section 3 we give our theoretical results. The main result, Theorem 1, gives an explicit representation of the stationary distribution of the colony size as the number of days becomes large. To our knowledge such a concise construction for a marked renewal process is not noted elsewhere in the literature. Moreover, when the distribution of the time between the batches is nonrandom, which is a typically case for a bee colony, the representation also enables us to obtain the characteristic function of the stationary distribution of the colony without relying on the renewal technique. See Corollary 2. In Theorem 3 we give results on the stationary distribution for the cyclic model. The practical implications of our results are that the colony size will not endlessly increase but will become stable. In effect it gives a baseline for the colony size. Moreover, our results give an explanation for the recovery in the colony size after swarming. If we suppose that the effect of environmental factors or disease is to either reduce the batch size and the hatching probability, increase mortality through reducing bee lifetime by shifting the lifetime distribution to the left, increase the time between batches or a combination of these effects then our results allow us to quantify their effect on the stationary distribution. In Section 4 we examine this for some parametric models. In Section 5 we give some simulations to verify the theoretical results and explore extinction probabilities as well as the time taken to attain the stationary distribution. The proofs of the results in Section 3 are given in the appendices.

## 2. Model and Notation

Let  $\tau_i$  be the time between batch  $i - 1$  and batch  $i$ ,  $\zeta_i$  be the number of eggs in batch  $i$ ,  $I_{ij} \in \{0, 1\}$  be the indicator that the  $j$ th egg in batch  $i$  hatches

and  $\eta_{ij}$  be the lifetime of bee  $j$  in batch  $i$ ,  $i \in \mathbb{N}$ . Then  $S_i = \sum_{j=1}^i \tau_i$  is the time that the  $i$ th batch of eggs is laid. The number of batches laid by time  $t$  is  $N_t = \sum_{i=1}^{\infty} \mathbf{1}_{[S_i, \infty)}(t)$ , where  $\mathbf{1}_A(\cdot)$  is the usual indicator function of the set  $A$ . Clearly, the number of bees that hatch in batch  $i$  is  $\xi_i = \sum_{j=1}^{\zeta_i} I_{ij}$ . Without loss of generality, shifting  $\{S_i\}$  by a fixed constant if necessary, we assume that the hatch time from an egg to a bee is zero so that the number of bees in the colony at time  $t$  is  $M_t = \sum_{i=1}^{N_t} \sum_{j=1}^{\xi_i} \mathbf{1}_{[S_i, S_i + \eta_{ij}]}(t) = \sum_{i=1}^{N_t} \sum_{j=1}^{\xi_i} \mathbf{1}_{\{S_i + \eta_{ij} \geq t\}}$ . In this expository work we suppose that  $\{\tau_i, i \in \mathbb{N}\}$  are independent and identically distributed (iid) as are  $\{I_{ij}, i, j \in \mathbb{N}\}$  and  $\{\zeta_i, i \in \mathbb{N}\}$ . We further suppose that  $\{\eta_{ij} : i, j \in \mathbb{N}\}$  are iid and that the sequences  $\{\tau_i, i \in \mathbb{N}\}$ ,  $\{\zeta_i, i \in \mathbb{N}\}$ ,  $\{I_{ij}, i, j \in \mathbb{N}\}$  and  $\{\eta_{ij}, i, j \in \mathbb{N}\}$  are independent.

Under our assumptions the times  $\{S_i, i \in \mathbb{N}\}$  at which the batches are laid form a renewal process. We let  $F_\tau(x)$  and  $F_\eta(x)$ ,  $x \geq 0$ , denote the distribution functions of  $\tau_1$  and  $\eta_{11}$  respectively and we suppose that  $\mathbb{E}(\tau_1)$ ,  $\text{Var}(\tau_1)$ ,  $\mathbb{E}(\eta_{11})$ ,  $\text{Var}(\eta_{11})$  are all finite. Note that our zero time is arbitrary and in general will not be the time at which a batch of eggs was laid. That is, the time after we start observing until the first batch is laid is  $\tilde{\tau}_1 = \tau_1 - u$  for some  $u$  representing the (random) time after the last batch of eggs was laid that our observations begun.

We assume that  $\tau_1$  takes positive integer values with distribution  $f_j = \mathbb{P}(\tau_1 = j)$ ,  $j \geq 1$ , where  $\sum_{j=1}^{\infty} f_j = 1$ . If  $\mathbb{P}(\tau_1 = 0) > 0$ , then we can redefine the  $\zeta_i$ 's so that the new renewal process has inter-renewal times taking positive integer values. Let  $X_n$  be the time till the next renewal from time  $n$ , then  $\{X_n\}$  is a Markov chain with transition probabilities  $p_{ij} := \mathbb{P}(X_{n+1} = j | X_n = i) = f_j$  for  $i = 1, j \geq 1$ ;  $p_{ij} = 1$  for  $i \geq 2, j = i - 1$ ; and  $p_{ij} = 0$  for all other cases. Let  $d$  be the greatest common divisor of  $\{j : f_j > 0\}$ . It is well-known that if  $d = 1$ , then  $\{X_n\}$  is ergodic with stationary distribution

$$\pi_k := \frac{1}{\mu}(1 - F(k - 1)), \quad k \geq 1,$$

while for  $d > 1$ ,  $\{X_n\}$  is not ergodic.

For simplicity, we also assume that the lifetimes  $\eta_{ij}$ 's only take non-negative integer values. We use  $\stackrel{d}{=}$  to stand for two random variables being equal in distribution,  $\stackrel{d}{\rightarrow}$  for the convergence in distribution and  $\sim$  denotes “is distributed as”. We let  $r = \mathbb{P}(I_{11} = 1)$  be the probability an egg hatches and set  $q_k = \mathbb{P}(\eta_{11} = k)$ ,  $k \geq 0$ . Let  $\tilde{\tau}_1$  be an integer-valued random variable with distribution  $\{\pi_k, k \geq 1\}$  such that  $\tilde{\tau}_1$  is independent of  $\{\tau_i, \xi_i, \eta_{ij} : i, j \geq 1\}$ . Define  $\mathcal{S}_i = \tilde{\tau}_1 + \tau_2 + \cdots + \tau_i$ ,  $i \geq 1$ .

### 3. Results

To state the main result, we need a metric to quantify the speed of convergence of  $\{M_n\}$  to its stationary distribution. We define the total variation distance between two integer valued random variables  $Y_1, Y_2$  as

$$d_{TV}(Y_1, Y_2) = \sup_{B \subset \mathbb{Z}} |\mathbb{P}(Y_1 \in B) - \mathbb{P}(Y_2 \in B)|,$$

where  $\mathbb{Z}$  is the space of all integers.

**Theorem 1.** *With the setup in §2, let  $\mathcal{M} := \sum_{i=1}^{\infty} \sum_{j=1}^{\xi_i} \mathbf{1}_{\{\mathcal{S}_i \leq \eta_{ij}+1\}}$ . Assume  $d = 1$  and  $\mathbb{E}\zeta_1 < \infty$ . Then*

$$\begin{aligned} & d_{TV}(M_n, \mathcal{M}) \\ & \leq \begin{cases} r\mathbb{E}(\zeta_1)\mathbb{E}(\max\{\eta_{11} + 1 - n, 0\}) \leq O(n^{-1}), & \text{if } F_\tau \text{ is degenerate,} \\ O(n^{-1/2}), & \text{if } F_\tau \text{ is non-degenerate.} \end{cases} \end{aligned}$$

Moreover,  $\mathbb{E}\mathcal{M} = r\mathbb{E}(\zeta_1)(\mathbb{E}\eta_{11} + 1)/\mathbb{E}\tau_1$ ,

$$\begin{aligned} \text{Var}(\mathcal{M}) &= \mathbb{E}\mathcal{M} - (\mathbb{E}\mathcal{M})^2 + r^2(\text{Var}(\zeta_1) - \mathbb{E}\zeta_1)[\mathbb{E}(\eta_{11} \wedge \eta_{12}) + 1]/\mathbb{E}\tau_1 \\ &+ \frac{(r\mathbb{E}\zeta_1)^2}{\mathbb{E}\tau_1} \sum_{v=0}^{\infty} \left( 2 \sum_{i=1}^v H(i) + v + 1 \right) q_v^2 \\ &+ 2 \frac{(r\mathbb{E}\zeta_1)^2}{\mathbb{E}\tau_1} \sum_{0 \leq v_1 < v_2} \left( \sum_{i=1}^{v_1} H(i) + v_1 + 1 + \sum_{i=v_2-v_1}^{v_2} H(i) \right) q_{v_1} q_{v_2}(1) \end{aligned}$$

where  $H(n) := \mathbb{E}N_n$  satisfies the renewal equation

$$H(n) = F(n) + \sum_{k=1}^n H(n-k)f_k. \quad (2)$$

Theorem 1 not only tells us that the marked renewal process  $\{M_n\}$  stabilises as  $n \rightarrow \infty$  but also gives an explicit representation of the stationary distribution. It enables us to simulate the stationary distribution, i.e., the distribution of  $M_\infty$ , without having to simulate the process  $\{M_n\}$  for a long time.

It can be observed from Theorem 1 that when  $\tau_i$ 's are not random (i.e.  $F_\tau$  is degenerate) and the lifetimes  $\eta_{ij}$  are bounded by a constant  $K$  then  $M_n$  reaches stationarity for  $n \geq K + 1$ . For a bee colony it is reasonable to take  $K$  to be 80 days. An implication of this is that after swarming (where the size of the colony is typically halved) in the stable case that we consider here, the colony size will quickly recover to the stationary size. Thus, at least in the stable case, no other mechanism is required to explain the recovery of the colony size after swarming.

Let  $\text{Pn}(\lambda)$  denote the Poisson distribution with mean  $\lambda$ . When  $\tau_i$ 's are not random, we have the following corollary.

**Corollary 2.** *If  $F_\tau$  is degenerate so that we can assume  $\mathcal{S}_i$ ,  $i \geq 1$ , are constants, then the stationary distribution of  $\{M_n\}$  has the characteristic function*

$$\phi_{\mathcal{M}}(u) := \mathbb{E}e^{iu\mathcal{M}} = \prod_{i=1}^{\infty} \psi_{\zeta} \left( r(e^{iu} - 1)(1 - F_{\eta}(\mathcal{S}_i - 2)) + 1 \right), \quad (3)$$

where  $F_{\eta}$  is the distribution function of  $\eta_{11}$  and  $\psi_{\zeta}(s) := \mathbb{E}s^{\zeta_1}$ . In particular, if  $\zeta_1 \sim \text{Pn}(\lambda)$ , then  $\mathcal{M} \sim \text{Pn}(r\lambda \sum_{i=1}^{\infty} (1 - F_{\eta}(\mathcal{S}_i - 2)))$ .

If the minimum size of a viable colony is  $\nu$  then using Corollary 2 to approximate  $\mathbb{P}(M_n < \nu)$  gives an approximate lower bound on the probability of colony extinction at time  $n$ . This is a lower bound as the colony may have



become extinct before  $n$  (i.e.  $\mathbb{P}(\min_{k \leq n} M_k < \nu) \geq \mathbb{P}(M_n < \nu)$ ). The computation of the extinction probabilities requires derivation of the distribution of  $\min_{k \leq n} M_k$  which is beyond the present work and for now we address this through simulations in Section 5. See Figure 4.

To allow the distribution  $F_{\zeta,i}$  of  $\zeta_i$ , the hatch probability  $r_i$  of eggs in batch  $i$  and the distribution  $F_{\eta,i}$  of  $\eta_{ij}$ ,  $j \geq 1$ , to all depend on time  $i$ , we propose the periodic model  $\{M_{lD+i}, i = 1, \dots, D\}$ ,  $l \in \mathbb{N}$ , to mimic the behaviour over a cycle of  $D$  days, where  $D = 365$  days represents a year. That is we assume  $F_{\zeta,\ell} = F_{\zeta,i}$ ,  $r_\ell = r_i$  and  $F_{\eta,\ell} = F_{\eta,i}$  if  $\ell = i \pmod{D}$ .

**Theorem 3.** *Assume  $\tau_1 \equiv 1$ ,  $\mathbb{E}\zeta_i < \infty$  for all  $1 \leq i \leq D$ , and  $\mathbb{P}(\eta_{ij} \leq K) = 1$ , then  $(M_{lD+n})_{l \in \mathbb{N}}$  reaches stationarity for  $lD + n > K$ , and the characteristic function of the stationary distribution is*

$$\phi_{n+lD}(t) = \prod_{i=n+lD-K}^{n+lD} \psi_{\zeta,i} \left( r_i (e^{it} - 1) (1 - F_{\eta,i}(n+lD-i-1)) + 1 \right),$$

where  $\psi_{\zeta,i}(s) := \mathbb{E}s^{\zeta_i}$ . In particular,  $\mathbb{E}M_{n+lD} = \sum_{i=n+lD-K}^{n+lD} (1 - F_{\eta,i}(n+lD-i-1))r_i\mathbb{E}\zeta_i$ , and if  $\zeta_i$  follows  $\text{Pn}(\lambda_i)$ , then  $M_{n+lD} \sim \text{Pn} \left( \sum_{i=n+lD-K}^{n+lD} r_i \lambda_i (1 - F_{\eta,i}(n+lD-i-1)) \right)$ .

This theorem establishes that the year to year behaviour of the colony size on a given day is stationary.

#### 4. Applications

We first use our analytic results to examine the effects of changing the distributions of the hatch probability, the batch size, the bee lifetime and the times between batches on the stationary distribution of the colony size. For example with a mean batch size of 1875, hatching probability 0.8, mean bee lifetime of 63 days and a constant time between batches of 1, then from Theorem 1 the

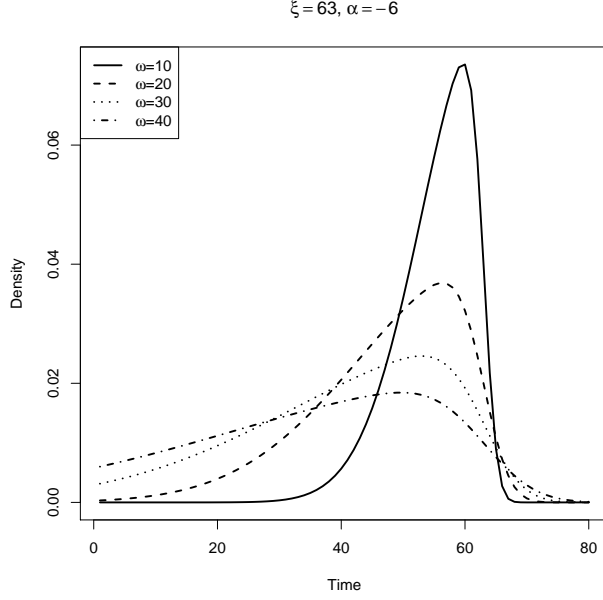


Figure 1: The  $SN(63, \omega^2, -6)$  density for  $\omega \in \{10, 20, 30, 40\}$ .

stationary mean colony size is 96,000. If the mean batch size drops to 1600 and the hatch probability drops to 0.75 while the mean bee lifetime remains at 63 days, the stationary mean colony size is now 76,800. If the mean bee lifetime also drops to 50, the stationary mean colony size is 61,200. Thus if the effect of a factor is to reduce the nectar flow into the hive reducing the mean batch size our results allow us to quantify this effect on the mean colony size. A reduction in the nectar flow could occur from a change in land use, change in climate, or a reduction in the numbers of forager bees. Alternatively, a disease may directly reduce the batch size or the hatching probability of the eggs.

To examine the behaviour of the colony size in more detail we model the lifetimes using a discretization of the skew normal distribution (Azzalini, 2013) truncated at zero as this distribution allows left skewed lifetime distributions. Left skewness is important as mortality will be low until bees start foraging and this may not occur until they are aged 40 days or so. The skew normal is a

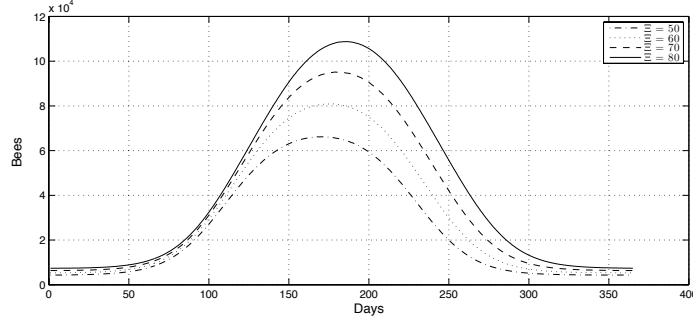


Figure 2: The mean of the daily colony size  $M_n$  when the lifetime  $\eta \sim SN(\Xi, 10^2, -6)$ , for the location parameter  $\Xi \in \{50, 60, 70, 80\}$  and a cyclic mean number of eggs given by  $\mathbb{E}(\xi_i) = r_i \mathbb{E}(\zeta_i) = 1500 \times \text{seas}(i) + 100$ .

three parameter distribution with a location parameter  $\Xi$ , a scale parameter  $\omega$  and a slant parameter  $\alpha$  (Azzalini, 2013). Its density is

$$f(x; \Xi, \omega, \alpha) = \frac{2}{\omega} \phi\left(\frac{x - \Xi}{\omega}\right) \Phi\left(\alpha \frac{x - \Xi}{\omega}\right),$$

where  $\phi$  and  $\Phi$  respectively denotes the probability density function and the cumulative distribution function of the standard normal law. Note that when  $\omega = 0$  the distribution of  $\eta$  is not random but constant equal to  $\Xi$ . We denote this density  $SN(\Xi, \omega^2, \alpha)$ . The mean of the skew normal distribution is  $\mu = \Xi + \omega \delta \sqrt{2/\pi}$  where  $\delta = \alpha / \sqrt{1 + \alpha^2}$ . If  $\alpha < 0$  then this distribution is skewed to the left. We discretise this distribution by rounding up to the nearest integer, with negative values rounded up to one.

In Figure 1 as an illustration we plot the  $SN(63, \omega^2, -6)$  density for varying values of  $\omega$ . This indicates that the shape of the distribution can become quite skewed. This skewness gives one way of representing an increase in bee mortality, with an increasingly heavy left tail when more bees die at a younger age. Varying the location and slant parameters further change the shape and location of this distribution, which gives the model considerable flexibility. Therefore we use this model to illustrate our results.

For convenience, we assume that we start observations on the first day of the year so that the subindex  $i$  represents the  $i$ th batch as well as the  $i$ th day (replace  $i$  with  $i \bmod 365$  if  $i > 365$ ) of the year.

In Figure 2 we use Theorem 3 and the skew-normal model for the lifetimes to investigate the effect of increasing the location parameter  $\Xi$  when there is a cyclic model for numbers of eggs laid. That is,  $\mathbb{E}(\xi_i) = r_i \mathbb{E}(\zeta_i) = 1500 \times \text{seas}(i) + 100$  where  $\text{seas}$  is given by the formula of Schmickl and Crailsheim (2007, p. 3). We suppose a Poisson distribution for  $\zeta_i$  and have noted that by the thinning property of the Poisson distribution (Grimmett and Stirzaker, 2001, p. 255 and p. 287),  $\xi_i$  also follows a Poisson distribution with mean  $\mathbb{E}(\xi_i) = r_i \mathbb{E}(\zeta_i)$ . This shows that increasing the location parameter  $\Xi$  increases the colony size.

In Figure D.6 in Appendix D we again use the skew-normal model for the lifetimes and in these figures we examine the effect of varying the other parameters in the lifetime distribution on the means as well as the seasonal effects. In Figure D.6 a) we see that increasing  $\omega$  decreases the mean seasonal population size. This is consistent with Figure 1 where we saw that increasing  $\omega$  resulted in a heavier left tail to the lifetime distribution. Figure D.6 b) illustrates the effect of changing values of  $\alpha$  and in Figure D.6 c), we plot the mean number of bees over a year with several distributions for  $\zeta_i$  and varying hatch probabilities  $r_i$ . As one would expect the mean population size increased as the mean number of eggs laid increases.

## 5. Simulations

To verify the theoretical results we conduct a series of simulations. We initially suppose  $\zeta_i \sim \text{Pn}(\mu_\zeta)$  and as the probability of an egg hatching is  $r$  we have  $\xi_i \sim \text{Pn}(\mu_\xi)$  with  $\mu_\xi = r\mu_\zeta$ . In Figure 3 we report ten simulations of our model with  $\eta \sim \text{SN}(63, 30^2, -6)$ , and  $\mathbb{E}(\xi_i) = r_i \mathbb{E}(\zeta_i) = 1500 \times \text{seas}(i) + 127$

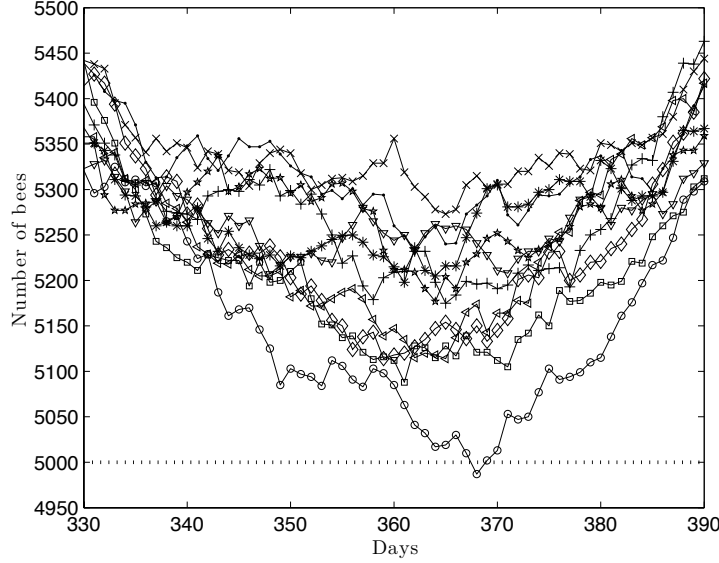


Figure 3: The number of bees during the first winter over ten different simulation, when the lifetime  $\eta \sim SN(63, 30^2, -6)$ , and  $\{\text{the number of hatched bees } \xi_i \sim \text{Pn}(1500 \times \text{seas}(i) + 127)\}$ . The dotted line is the threshold of 5000 under which the colony collapses.

during the first winter. We can see that one of the simulations resulted in the number of bees being under the nominal threshold of 5000 and in that case the colony would have become extinct. To examine this further if the laying times are constants and the distribution of the numbers of eggs laid, the hatching probability and the distribution of the bee lifetimes are specified, we can apply Corollary 2 to determine lower bounds on the extinction probabilities. In general, we can always use simulations to estimate extinction probabilities, although for many parameter values this will be small. In Figure 4 we plot the extinction probability (supposing that extinction occur when the number of bees in a colony is less than 5000) according to the mean number of eggs laid per day. Here, we simulated the probability of extinction over 30 days as a function of the mean number of eggs laid. This number corresponds to mid-winter in our model. Thus our model may be used to determine the effects of changes in the underlying parameters on the extinction probabilities. Note that

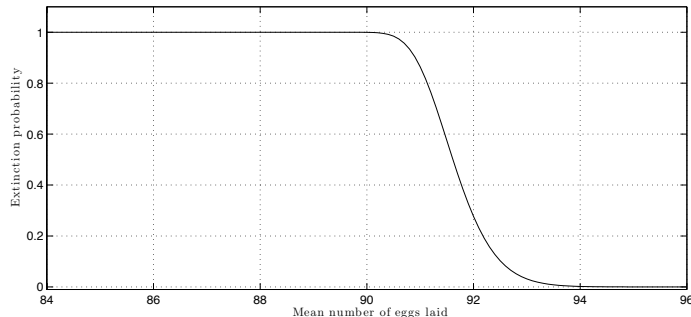


Figure 4: The simulated extinction probability as a function of the mean number of eggs laid per day obtained in Corollary 2. Here we took  $\eta \sim SN(63, 10^2, -6)$  and  $\xi_i \sim \text{Pn}(L)$  with  $L \in [80; 95]$ .

it only requires a small change in the mean number of eggs laid per day for the extinction probability to increase from zero to one.

In Figure 5 we plot the number of bees in the colony during a period slightly longer than one year using the model described above but incorporating swarming. That is we assumed that when the colony size was more than 80000, half of the colony leave the colony. Clearly, in the right conditions the colony size recovers and as noted above no other mechanism is required to explain the recovery of the colony size.

A further quantity of interest is the time until the colony size attains the stationary distribution. In our case, as the maximum lifetime of a bee is bounded by 80, the colony size attains the stationary distribution after no more than 81 days as long as the number of eggs laid is sufficiently large. This can be observed in Figure 5. Indeed, 80 days after the colony splits, the number of bees in the hive is as if swarming never happened. From an evolutionary perspective by splitting into two, colonies that swarm have an increased probability that one the colonies will survive.

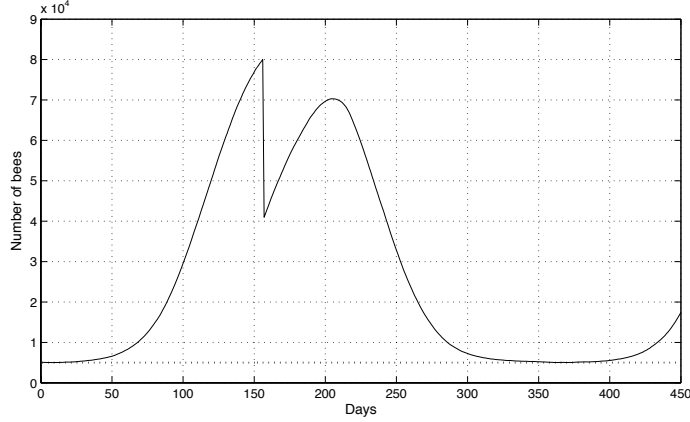


Figure 5: The number of bees simulated over 450 days according to the cyclic model with swarming. Here  $\eta \sim SN(63, 10^2, -6)$ , and  $\xi_i \sim Pn(1500 \times seas(i) + 90)$  and we allowed swarming when the number of bees exceeded 80000.

## 6. Discussion

Our model is an encouraging first step in modelling the dynamics of a bee colony. We have developed a probabilistic model for the dynamics of a bee colony in an idealised bee world, which may not of course be attained. This degree of simplicity was necessary to obtain analytic results, but nevertheless, our results explain why a typical colony size in summer is around 90,000 bees. Our interpretation is that the stationary distribution represents a state attainable under ideal conditions and thus in many ways is an upper bound. We extended the stationarity results to a more naturalistic seasonal model and showed the marginal distribution of the colony size on a given stage of the season was stationary. Our model also allowed us to give a lower bound for extinction probabilities and demonstrate how the colony size recovers after swarming, as seen in the natural system. The simple expression for the mean colony size is easily computed in terms of the hatching probability and the mean of the numbers of eggs laid, the expected bee lifetime and the expected time between the laying of batches of eggs. A determination of how a factor such as environmental

perturbations, disease pressure or weather impacts the colony size (for example through increased mortality or reduced nectar flow and hence the number of eggs laid per day) then allows an examination of its effect on colony size.

There are several ways in which our work can be extended. An extension of our results would first be to consider the joint behaviour of the colony size over several time points which would potentially allow the derivation of exact extinction probabilities in addition to the lower bound. That is to consider the colony size  $\{M_{lD+i}, i = 1, \dots, D\}$  over a year as “functional data” (Ramsay and Silverman, 1997) and examine the behaviour of this function. Corollary 2 examines the marginal distributions of this quantity. One could then consider the effects of long term trends on the colony size and survival. In particular conditional probabilities of survival given a particular colony size could in theory be derived. That is, given the current colony size of say 7000 bees at a given time of year under what conditions would the probability of surviving over winter be 90% for example? Such calculations could indicate whether supplementary feeding was required to increase the number of eggs laid per day to help ensure survival. By relating the effects of environmental changes and disease to these quantities, their effect on colony size may also be examined. Whilst our current results are for stable and cyclic effects we anticipate extending these to even more complex and realistic situations in the future. For example, the availability of nectar in the hive may depend on the numbers of forager bees so that the average batch size could be taken to be dependent on the number of forager bees. In their deterministic models, Khoury et al. (2013) give the steady state population as a function of bee death rate and food collection rate. Bee death rate relates to our bee lifetime distribution. However, we can consider more subtle effects. We do not explicitly include food collection rate in the current model but we have argued this could be included in the models for the mean batch size. The availability of nectar could also depend on the age related efficiency of the forager bees and the average batch size could be taken to be dependent on the ages of the forager bees. These more complex feedback models



are technically challenging and beyond the scope of the current work.

Regarding more technical extensions, recall that a lattice distribution is the distribution of a random variable taking possible values of the form  $\{a + bk, a, b \text{ are constants}, b \neq 0, k \in \mathbb{Z}\}$ . The distribution  $F_\tau$  of the renewal times  $\{\tau_i\}$  in Theorem 1 is a special case of a lattice distribution with  $b = 1$ . One can easily adapt the proof of Theorem 1 to obtain the stationary distribution when  $F_\tau$  either follows a general lattice distribution or a non-lattice distribution. We also note that one can also study  $\{M_n\}$  as a discrete-time queueing system with batch arrivals (see Takagi (1993)).

We conclude that any change in the parameters of the lifetime distribution or the numbers of eggs laid can affect the population size. This allows a variety of ways in which a disease for example can affect the lifetime distribution and hence colony size. Whilst their properties can be difficult to derive, we believe that probabilistic models for the size of a bee colony are mathematically tractable and provide useful and readily interpretable information on the dependence of the colony size on the times between batches of eggs being laid, the sizes of the batches, the hatching probability and bee lifetime. We further believe they and their extensions have considerable potential both in understanding the behaviour of bee colonies and also provide new classes of stochastic models worthy of further research.

## References

- Al-Ghamdi, A. A., H. F. Abou-Shaara, and A. A. Mohamed (2014). Hatching rates and some characteristics of yemeni and carniolan honey bee eggs. *Journal of Entomology and Zoology Studies* 2(1), 6–10.
- Arundel, J. (2011). Understanding the spread of honeybee pests and diseases:

- An agent-based modelling approach. Technical report, Austrian Government Rural Industries Research and Development Corporation.
- Azzalini, A. (2013). *The Skew-Normal and Related Families*. Cambridge University Press. Cambridge Books Online.
- Barbour, A. D., L. Holst, and S. Janson (1992). *Poisson Approximation*. Oxford Univ. Press.
- Becher, M. A., V. Grimm, P. Thorbek, J. Horn, P. J. Kennedy, and J. L. Osborne (2014). Beehave: a systems model of honey bee colony dynamics and foraging to explore multifactorial causes of colony failure. *Journal of Applied Ecology* 51, 470–482.
- Becher, M. A., J. L. Osborne, P. Thorbek, P. J. Kennedy, and V. Grimm (2013). Towards a systems approach for understanding honey bee decline: a stocktaking and synthesis of existing models. *The Journal of Applied Ecology* 50(4), 868–880.
- Bishop, J., H. E. Jones, M. Lukac, and S. G. Potts (2016). Insect pollination reduces yield loss following heat stress in faba bean (*vicia faba* l.). *Agriculture, Ecosystems & Environment* 220, 89–96.
- Blaauw, B. R. and R. Isaacs (2014). Flower plantings increase wild bee abundance and the pollination services provided to a pollination-dependent crop. *Journal of Applied Ecology* 51(4), 890–898.
- Botias, C., A. David, J. Horwood, A. Abdul-Sada, E. Nicholls, E. Hill, and D. Goulson (2015). Neonicotinoid residues in wildflowers, a potential route of chronic exposure for bees. *Environ. Sci. Technol.* 49 (21), 12731–12740.
- Breeze, T. D., S. P. Roberts, and S. G. Potts (2012). The decline of england’s bees. Technical report, University of Reading Centre for Agri-Environment Research on commission from Friends of the Earth England.

- Bull, J. C., E. V. Ryabov, G. Prince, A. Mead, C. Zhang, L. A. Baxter, J. K. Pell, J. L. Osborne, and D. Chandler (2012). A strong immune response in young adult honeybees masks their increased susceptibility to infection compared to older bees. *PLoS Pathog* 8(12), e1003083. doi:10.1371/journal.ppat.1003083.
- Capri, E., M. Higes, and K. M. Kasiotis (2013). Bee health in europe - facts & figures 2013. Technical report, OPERA Research Centre, Universit Cattolica del Sacro Cuore.
- Chandler, D., G. Davidson, J. K. Pell, V. B. Shaw, and K. D. Sunderland (2000). Fungal biocontrol of acari. *Biocontrol Science and Technology* 10, 357–384.
- Chauzat, M.-P., M. Laurent, M.-P. Riviere, C. Saugeon, P. Hendrikx, and M. R.-C. on behalf of the EPILOBEE consortium (2014). A pan-european epidemiological study on honey bee colony losses 2012-2013. Technical report, EPILOBEE.
- Chen, L. H. Y., L. Goldstein, and Q. Shao (2011). *Normal approximation by Stein's method*. Springer-Verlag.
- Collier, R. A. (2009). Identify reasons why food security may be an issue requiring specific attention. *DEFRA project FO0416*.
- Collins, A. M. (2004). Variation in time of egg hatch by the honey bee, *apis mellifera* (hymenoptera: Apidae). *Annals of the Entomological Society of America* 97(1), 140–146.
- Cook, D. C., M. B. Thomas, S. A. Cunningham, D. I. Anderson, and P. J. D. Barro (2007). Predicting the economic impact of an invasive species on an ecosystem service. *Ecological Applications* 17 (6), 1832–1840.
- Datta, S., J. C. Bull, G. E. Budge, and M. J. Keeling (2013). Modelling the spread of american foulbrood in honey bees. *Journal of The Royal Society Interface* 10(88), DOI: 10.1098/rsif.2013.0650.

- Dively, G. P., M. S. Embrey, A. Kamel, D. J. Hawthorne, and J. S. Pettis (2015). Assessment of chronic sublethal effects of imidacloprid on honey bee colony health. *PLoS ONE* 10(3), e0118748. doi:10.1371/journal.pone.0118748.
- Fürst, M. A., D. P. McMahon, J. L. Osborne, R. J. Paxton, and M. J. F. Brown (2014). Disease associations between honeybees and bumblebees as a threat to wild pollinators. *Nature* 506, 364–366.
- Godfray, H. C. J., T. Blacqui re, L. M. Field, R. S. Hails, G. Petrokofsky, S. G. Potts, N. E. Raine, A. J. Vanbergen, and A. R. McLean (2014). A restatement of the natural science evidence base concerning neonicotinoid insecticides and insect pollinators. *Proc. R. Soc. B* 281, 20140558.
- Goodman, R. and P. Kaczynski (2015). *The Australian Beekeeping Guide*. Rural Industries Research and Development Corporation.
- Gordon, R., N. Bresolin-Schott, and I. East (2014). Nomadic beekeeper movements create the potential for widespread disease in the honeybee industry. *The Journal of the Australian Veterinary Association Ltd.* 92, 283–290.
- Grimmett, G. and D. Stirzaker (2001). *Probability and Random Processes* (Third ed.). Oxford University Press.
- Hafi, A., N. Millist, K. Morey, P. Caley, and B. Buetre (2012). A benefit-cost framework for responding to an incursion of varroa destructor. Technical report, Australian Bureau of Agricultural and Resource Economics and Sciences.
- Khoury, D. S., A. B. Barron, and M. R. Myerscough (2013). Modelling food and population dynamics in honey bee colonies. *PLOS ONE* 8, DOI: 10.1371/journal.pone.0059084.
- Lawler, G. F. (2006). *Introduction to Stochastic Processes* (Second ed.). Chapman & Hall.

- Lever, J. J., E. H. Nes, M. Scheffer, and J. Bascompte (2014). The sudden collapse of pollinator communities. *Ecology Letters* 17(3), 350–359.
- Lindvall, T. (1992). *Lectures on the coupling method*. John Wiley & Sons.
- Lonsdorf, E., C. Kremen, T. Ricketts, R. Winfree, and N. W. S. Greenleaf (2009). Modelling pollination services across agricultural landscapes. *Annals of Botany* 103, 1589–1600.
- Mattila, H. R., J. L. Harris, and G. W. Otis (2001). Timing of production of winter bees in honey bee (*Apis mellifera*) colonies. *Insectes sociaux* 48(2), 88–93.
- Owen, R. (2015). *The Australian Beekeeping Manual*. Exisle Publishing Pty Ltd.
- Perry, C. J., E. Svik, M. R. Myerscough, and A. B. Barron (2015). Rapid behavioral maturation accelerates failure of stressed honey bee colonies. *Proceedings of the National Academy of Sciences* 112(11), 3427–3432.
- Pettis, J. S., E. M. Lichtenberg, M. Andree, J. Stitzinger, R. Rose, and D. Vanengelsdorp (2013). Crop pollination exposes honey bees to pesticides which alters their susceptibility to the gut pathogen *Nosema ceranae*. *PLoS ONE* 8(7), e70182. doi: 10.1371/journal.pone.0070182.
- Potts, S. G., S. P. M. Roberts, R. Dean, G. Marris, M. Brown, R. Jones, and J. Settele (2009). Declines of managed honey bees and beekeepers in Europe. *Journal of Apicultural Research* 49(1), 15–22.
- Rader, R., I. Bartomeus, L. A. Garibaldi, M. P. D. Garratt, B. G. Howlett, R. Winfree, S. A. Cunningham, M. M. Mayfield, A. D. Arthur, G. K. S. Andersson, R. Bommarco, C. Brittain, L. G. Carvalheiro, N. P. Chacoff, M. H. Entling, B. Foully, B. M. Freitas, B. Gemmill-Herren, J. Ghazoul, S. R. Griffin, C. L. Gross, L. Herbertsson, F. Herzog, J. Hipólito, S. Jaggar, F. Jauker, A.-M. Klein, D. Kleijn, S. Krishnan, C. Q. Lemos, S. A. M.

- Lindström, Y. Mandelik, V. M. Monteiro, W. Nelson, L. Nilsson, D. E. Pattemore, N. de O. Pereira, G. Pisanty, S. G. Potts, M. Reemer, M. Rundlöf, C. S. Sheffield, J. Scheper, C. Schüepp, H. G. Smith, D. A. Stanley, J. C. Stout, H. Szentgyörgyi, H. Taki, C. H. Vergara, B. F. Viana, and M. Woyciechowski (2016). Non-bee insects are important contributors to global crop pollination. *Proceedings of the National Academy of Sciences* 113, 146–151.
- Ramsay, J. and B. Silverman (1997). *Functional Data Analysis*. Springer-Verlag New York.
- Ryabov, E. V., G. R. Wood, J. M. Fannon, J. D. Moore, J. C. Bull, D. Chandler, A. Mead, N. Burroughs, and D. J. Evans (2014). A virulent strain of deformed wing virus (dwv) of honeybees (*apis mellifera*) prevails after varroa destructor-mediated, or in vitro, transmission. *PLOS Pathog* 10(6), DOI: 10.1371/journal.ppat.1004230.
- Schmickl, T. and K. Crailsheim (2007). HoPoMo: A model of honeybee intracolony population dynamics and resource management. *Ecological Modelling* 204, 219–245.
- Schulz, D. J., Z.-Y. Huang, and G. E. Robinson (1998). Effects of colony food shortage on behavioral development in honey bees. *Behavioral Ecology and Sociobiology* 42(5), 295–303.
- Takagi, H. (1993). *Queueing Analysis: Discrete-time Systems*. Vol. 3, Elsevier Science Publishers.
- UN (2015). World population prospects: The 2015 revision, key findings and advance tables. working paper no. esa/p/wp.241. Technical report, United Nations Department of Economic and Social Affairs Population Division.
- Vanbergen, A. J., M. S. Heard, T. Breeze, S. G. Potts, and N. Hanley (2014). Status and value of pollinators and pollination services: A report to the department for environment, food and rural affairs (defra). Technical report, DEFRA.

## Appendix A. Proof of Theorem 1

First of all, since  $\xi_1 = \sum_{j=1}^{\zeta_1} I_{1j}$  is a mixed binomial, we have

$$\mathbb{E}\xi_1 = \mathbb{E} \left\{ \mathbb{E} \left( \sum_{j=1}^{\zeta_1} I_{1j} \middle| \zeta_1 \right) \right\} = \mathbb{E}(r\zeta_1) = r\mathbb{E}\zeta_1 \quad (\text{A.1})$$

and

$$\text{Var}(\xi_1) = \mathbb{E} \left\{ \text{Var} \left( \sum_{j=1}^{\zeta_1} I_{1j} \middle| \zeta_1 \right) \right\} + \text{Var} \left\{ \mathbb{E} \left( \sum_{j=1}^{\zeta_1} I_{1j} \middle| \zeta_1 \right) \right\} = r(1-r)\mathbb{E}\zeta_1 + r^2\text{Var}(\zeta_1). \quad (\text{A.2})$$

If  $F$  is non-degenerate, we write  $T_n = \tau_2 + \dots + \tau_n$ , then the condition  $d = 1$  implies  $d_{TV}(T_n, T_n + 1) = O(n^{-1/2}) \rightarrow 0$  as  $n \rightarrow \infty$  (Lindvall, 1992, pp. 41–43). This ensures

$$\begin{aligned} d_{TV}(\mathcal{S}_n, S_n) &\leq d_{TV}(\mathcal{S}_n, T_n) + d_{TV}(T_n, S_n) \\ &\leq (\mathbb{E}\tilde{\tau}_1 + \mathbb{E}\tau_1)d_{TV}(T_n, T_n + 1) = O(n^{-1/2}) \rightarrow 0 \end{aligned} \quad (\text{A.3})$$

as  $n \rightarrow \infty$ . If  $F_\tau$  is degenerate, then  $\mathcal{S}_n = S_n$ , giving  $d_{TV}(\mathcal{S}_n, S_n) = 0$ . Hence, we obtain

$$d_{TV}(\mathcal{S}_n, S_n) = \begin{cases} 0, & \text{if } F_\tau \text{ is degenerate,} \\ O(n^{-1/2}), & \text{if } F_\tau \text{ is non-degenerate.} \end{cases} \quad (\text{A.4})$$

Let  $\mathcal{N}_n := \sum_{i=1}^{\infty} \mathbf{1}_{\{\mathcal{S}_i \leq n\}}$  and  $\mathcal{M}_n := \sum_{i=1}^{\mathcal{N}_n} \sum_{j=1}^{\xi_i} \mathbf{1}_{\{\mathcal{S}_i + \eta_{ij} \geq n\}}$ , we first establish that

$$d_{TV}(M_n, \mathcal{M}_n) = \begin{cases} 0, & \text{if } F_\tau \text{ is degenerate,} \\ O(n^{-1/2}), & \text{if } F_\tau \text{ is non-degenerate.} \end{cases} \quad (\text{A.5})$$

In fact, when  $F_\tau$  is degenerate, we have  $M_n \stackrel{d}{=} \mathcal{M}_n$ , so (A.5) is obvious. For the case that  $F_\tau$  is non-degenerate, we have  $\sigma_\tau := \sqrt{\text{Var}(\tau_1)} \in (0, \infty)$ . Take  $a_n = \lfloor 0.75n \rfloor$ ,  $b_n = \lfloor 0.5n/\mathbb{E}\tau_1 \rfloor$ , where  $\lfloor x \rfloor$  is the biggest integer less than or

equal to  $x$ . By (Barbour et al., 1992, pp. 253–254), enlarging the probability space if necessary, we can construct iid  $\{\tau'_i\}$  having the same distribution as that of  $\tau_1$  and independent of  $\{\xi_i, \eta_{ij} : i, j \in \mathbb{N}\}$  such that  $\left\{S'_k := \sum_{j=1}^k \tau'_j : k \in \mathbb{N}\right\}$  satisfies

$$\mathbb{P}(S'_{b_n} \neq \mathcal{S}_{b_n}) = d_{TV}(S_{b_n}, \mathcal{S}_{b_n}) = O(b_n^{-1/2}) = O(n^{-1/2}), \quad (\text{A.6})$$

where the second equality is due to (A.4). Set  $\mathcal{S}'_j := \begin{cases} \mathcal{S}_j, & \text{if } j \leq b_n, \\ \mathcal{S}_{b_n} + \sum_{l=b_n+1}^j \tau'_l, & \text{if } j > b_n, \end{cases}$   
 $N'_n := \sum_{i=1}^{\infty} \mathbf{1}_{\{S'_i \leq n\}}$ ,  $M'_n := \sum_{i=1}^{N'_n} \sum_{j=1}^{\xi_i} \mathbf{1}_{\{S'_i + \eta_{ij} \geq n\}}$ ;  $\mathcal{N}'_n := \sum_{i=1}^{\infty} \mathbf{1}_{\{\mathcal{S}'_i \leq n\}}$ ,  
 $\mathcal{M}'_n := \sum_{i=1}^{\mathcal{N}'_n} \sum_{j=1}^{\xi_i} \mathbf{1}_{\{\mathcal{S}'_i + \eta_{ij} \geq n\}}$ , then  $M'_n \stackrel{d}{=} M_n$ ,  $\mathcal{M}'_n \stackrel{d}{=} \mathcal{M}_n$  and

$$d_{TV}(M_n, \mathcal{M}_n) \leq \mathbb{P}(M'_n \neq \mathcal{M}'_n).$$

Therefore, it suffices to show that  $\mathbb{P}(M'_n \neq \mathcal{M}'_n) = O(n^{-1/2})$ . To this end, we observe that

$$\{M'_n \neq \mathcal{M}'_n\} \subset \{S'_{b_n} \neq \mathcal{S}_{b_n}\} \cup \{S'_{b_n} > a_n\} \cup \{S'_{b_n} = \mathcal{S}_{b_n} \leq a_n, M'_n \neq \mathcal{M}'_n\}. \quad (\text{A.7})$$

For the event  $\{S'_{b_n} \neq \mathcal{S}_{b_n}\}$ , (A.6) ensures that  $\mathbb{P}(S'_{b_n} \neq \mathcal{S}_{b_n}) = O(n^{-1/2})$ . For the event  $\{S'_{b_n} > a_n\}$ , we have

$$\mathbb{P}(S'_{b_n} > a_n) = \mathbb{P}\left(\frac{S'_{b_n} - \mathbb{E}S'_{b_n}}{\sqrt{\text{Var}(S'_{b_n})}} > c_n \sqrt{n}\right) = O(n^{-1/2}) + \int_{c_n \sqrt{n}}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = O(n^{-1/2}),$$

where  $c_n \asymp \frac{\sqrt{2\mathbb{E}\tau_1}}{4\sigma_\tau}$ , the second equality is due to (Chen et al., 2011, Theorem 11.2) and the last equality follows from (Chen et al., 2011, (2.82)). Therefore, it remains to show that the final event of (A.7) has probability of order no more than  $O(n^{-1/2})$ . This can be worked out as follows. On  $\{S'_{b_n} = \mathcal{S}_{b_n} \leq a_n, M'_n \neq \mathcal{M}'_n\}$ , the two marked renewal processes coincide from time  $a_n$ , so  $M'_n \neq \mathcal{M}'_n$  can only happen when at least one of the bees before  $a_n$  are still alive at time  $n$ . That is, the bee must have lived through the interval  $[a_n, n]$  with lifetime  $\eta_{ij} \geq n - a_n$ , giving

$$\{S'_{b_n} = \mathcal{S}_{b_n} \leq a_n, M'_n \neq \mathcal{M}'_n\} \subset \cup_{i=1}^{b_n-1} \cup_{j=1}^{\xi_i} \{\eta_{ij} \geq n - a_n\}.$$



This yields

$$\begin{aligned}
& \mathbb{P}(S'_{b_n} = \mathcal{S}_{b_n} \leq a_n, M'_n \neq \mathcal{M}'_n) \\
& \leq (b_n - 1) \mathbb{E} \xi_1 \mathbb{P}(\eta_{11} \geq n - a_n) \leq b_n \mathbb{E} \xi_1 \int_{n-a_n}^{\infty} \frac{x^2}{(n-a_n)^2} dF_{\eta_{11}}(x) \\
& \leq \frac{b_n r \mathbb{E} \zeta_1 \mathbb{E} \eta_{11}^2}{(n-a_n)^2} = O(n^{-1}),
\end{aligned}$$

where the last inequality is due to (A.1). This completes the proof of (A.5).

Set  $\mathcal{X}_n := \mathcal{S}_{\mathcal{N}_n+1} - n$ ,  $\mathcal{B}_n := n - \mathcal{S}_{\mathcal{N}_n}$ , then  $\mathcal{X}_n$  is the time to the next renewal from time  $n$  and  $\mathcal{B}_n$  is the time from the renewal at or before time  $n$  to time  $n$ . Clearly  $\mathcal{X}_0 = \tilde{\tau}_1 \sim \{\pi_k\}$ . For  $l \in \mathbb{N}$ ,

$$\begin{aligned}
\mathbb{P}(\mathcal{X}_1 = l) &= \mathbb{P}(\mathcal{X}_1 = l | \tilde{\tau}_1 = 1) \mathbb{P}(\tilde{\tau}_1 = 1) + \mathbb{P}(\mathcal{X}_1 = l | \tilde{\tau}_1 = l+1) \mathbb{P}(\tilde{\tau}_1 = l+1) \\
&= f_l \pi_1 + \pi_{l+1} = \pi_l,
\end{aligned}$$

hence  $\mathcal{X}_1 \sim \{\pi_k\}$ . By the induction on  $n$ ,  $\mathcal{X}_n$  has the distribution  $\{\pi_k\}$  for all  $n$ . Next, we establish that  $\mathcal{B}_n + 1$  follows the distribution  $\{\pi_k\}$  as well. To this end, for  $l \in \mathbb{N}$ ,

$$\mathbb{P}(\mathcal{B}_n \geq l) = \mathbb{P}(\mathcal{N}_n - \mathcal{N}_{n-l} = 0) = \mathbb{P}(\mathcal{X}_{n-l} \geq l+1) = \sum_{k \geq l+1} \pi_k,$$

so the claim follows immediately. A direct consequence of the claim is that

$$n+1 - (\mathcal{S}_{\mathcal{N}_n}, \mathcal{S}_{\mathcal{N}_n-1}, \mathcal{S}_1) \stackrel{d}{=} (\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_{\mathcal{N}_n}),$$

which ensures

$$\mathcal{M}_n \stackrel{d}{=} \sum_{i=1}^{\mathcal{N}_n} \sum_{j=1}^{\xi_i} \mathbf{1}_{\{n+1-\mathcal{S}_i+\eta_{ij} \geq n\}} = \sum_{i=1}^{\mathcal{N}_n} \sum_{j=1}^{\xi_i} \mathbf{1}_{\{1+\eta_{ij} \geq \mathcal{S}_i\}} =: \tilde{\mathcal{M}}_n.$$

Clearly,

$$d_{TV}(\mathcal{M}_n, \mathcal{M}) \leq \mathbb{P}(\tilde{\mathcal{M}}_n \neq \mathcal{M}) \leq \mathbb{E}(\mathcal{M} - \tilde{\mathcal{M}}_n) = \mathbb{E} \sum_{i=\mathcal{N}_n+1}^{\infty} \sum_{j=1}^{\xi_i} \mathbf{1}_{\{1+\eta_{ij} \geq \mathcal{S}_i\}}. \tag{A.8}$$

However,  $\mathcal{S}_{\mathcal{N}_n+l} \geq n+l$ , it follows from (A.8) that

$$\begin{aligned} d_{TV}(\mathcal{M}_n, \mathcal{M}) &\leq \mathbb{E} \sum_{i=n+1}^{\infty} \sum_{j=1}^{\xi_i} \mathbf{1}_{\{1+\eta_{ij} \geq i\}} \\ &= \mathbb{E} \xi_1 \mathbb{E}(\max\{\eta_{11} + 1 - n, 0\}) \end{aligned} \quad (\text{A.9})$$

$$\begin{aligned} &\leq \mathbb{E} \xi_1 \mathbb{E}(\eta_{11} \mathbf{1}_{\{\eta_{11} \geq n\}}) \leq n^{-1} \mathbb{E} \xi_1 \mathbb{E}(\eta_{11}^2) \\ &= O(n^{-1}). \end{aligned} \quad (\text{A.10})$$

Combining (A.5), (A.9), (A.10) and (A.1) gives

$$\begin{aligned} d_{TV}(M_n, \mathcal{M}) &\leq d_{TV}(M_n, \mathcal{M}_n) + d_{TV}(\mathcal{M}_n, \mathcal{M}) \\ &\leq \begin{cases} r\mathbb{E}(\zeta_1) \mathbb{E}(\max\{\eta_{11} + 1 - n, 0\}) \leq O(n^{-1}), & \text{if } F_\tau \text{ is degenerate,} \\ O(n^{-1/2}), & \text{if } F_\tau \text{ is non-degenerate.} \end{cases} \end{aligned}$$

For  $\mathbb{E}\mathcal{M}$ , we have from (A.1) that

$$\begin{aligned} \mathbb{E}\mathcal{M} &= \sum_{i=1}^{\infty} \mathbb{E} \left\{ \mathbb{E} \left( \sum_{j=1}^{\xi_i} \mathbf{1}_{\{\mathcal{S}_i \leq \eta_{ij}+1\}} \middle| \xi_i, \mathcal{S}_i \right) \right\} \\ &= \sum_{i=1}^{\infty} \mathbb{E} \{ \xi_i \mathbb{P}(\mathcal{S}_i \leq \eta_{11} + 1 | \mathcal{S}_i) \} \\ &= \mathbb{E}(\xi_1) \sum_{i=1}^{\infty} \sum_{l=0}^{\infty} \mathbb{P}(\mathcal{S}_i \leq l+1) q_l \\ &= r\mathbb{E}(\zeta_1) \sum_{l=0}^{\infty} \mathbb{E}\mathcal{N}_{l+1} q_l = r\mathbb{E}(\zeta_1) \sum_{l=0}^{\infty} \frac{l+1}{\mathbb{E}\tau_1} q_l \\ &= r\mathbb{E}(\zeta_1) (\mathbb{E}\eta_{11} + 1) / \mathbb{E}\tau_1. \end{aligned}$$

For  $\text{Var}(\mathcal{M})$ , we write  $\mathbb{P}(\eta_{11} \wedge \eta_{12} = u) = q'_u$ , then decomposing according to

$i = l, j = m; i = l, j \neq m; \text{ and } i \neq l$  for the second equation below, we obtain

$$\begin{aligned}
\mathbb{E}(\mathcal{M}^2) &= \mathbb{E} \sum_{i=1}^{\infty} \sum_{l=1}^{\infty} \sum_{j=1}^{\xi_i} \sum_{m=1}^{\xi_l} \mathbf{1}_{\{\mathcal{S}_i \leq \eta_{ij}+1, \mathcal{S}_l \leq \eta_{lm}+1\}} \\
&= \mathbb{E} \sum_{i=1}^{\infty} \sum_{j=1}^{\xi_i} \mathbf{1}_{\{\mathcal{S}_i \leq \eta_{ij}+1\}} + \mathbb{E} \sum_{i=1}^{\infty} \sum_{1 \leq j \neq m \leq \xi_i} \mathbf{1}_{\{\mathcal{S}_i \leq \eta_{ij} \wedge \eta_{im}+1\}} \\
&\quad + \mathbb{E} \sum_{1 \leq i \neq l < \infty} \sum_{j=1}^{\xi_i} \sum_{m=1}^{\xi_l} \mathbf{1}_{\{\mathcal{S}_i \leq \eta_{ij}+1, \mathcal{S}_l \leq \eta_{lm}+1\}} \\
&= \mathbb{E} \mathcal{M} + \sum_{i=1}^{\infty} \mathbb{E} \left\{ \mathbb{E} \left( \sum_{1 \leq j \neq m \leq \xi_i} \mathbf{1}_{\{\mathcal{S}_i \leq \eta_{ij} \wedge \eta_{im}+1\}} \middle| \mathcal{S}_i, \xi_i \right) \right\} \\
&\quad + \sum_{1 \leq i \neq l < \infty} \mathbb{E} \left\{ \mathbb{E} \left( \sum_{j=1}^{\xi_i} \sum_{m=1}^{\xi_l} \mathbf{1}_{\{\mathcal{S}_i \leq \eta_{ij}+1, \mathcal{S}_l \leq \eta_{lm}+1\}} \middle| \mathcal{S}_i, \mathcal{S}_l, \xi_i, \xi_l \right) \right\},
\end{aligned}$$

which can be simplified to

$$\begin{aligned}
\mathbb{E}(\mathcal{M}^2) &= \mathbb{E} \mathcal{M} + \sum_{i=1}^{\infty} \sum_{u=0}^{\infty} q'_u \mathbb{E} \{ \xi_i (\xi_i - 1) \mathbb{P}(\mathcal{S}_i \leq u+1 | \mathcal{S}_i) \} \\
&\quad + \sum_{1 \leq i \neq l < \infty} \sum_{u_1=0}^{\infty} \sum_{u_2=0}^{\infty} q_{u_1} q_{u_2} \mathbb{E} \{ \xi_i \xi_l \mathbb{P}(\mathcal{S}_i \leq u_1+1, \mathcal{S}_l \leq u_2+1 | \mathcal{S}_i, \mathcal{S}_l) \} \\
&= \mathbb{E} \mathcal{M} + \mathbb{E}(\xi_1^2 - \xi_1) \sum_{i=1}^{\infty} \sum_{u=0}^{\infty} q'_u \mathbb{P}(\mathcal{S}_i \leq u+1) \\
&\quad + (\mathbb{E} \xi_1)^2 \sum_{1 \leq i \neq l < \infty} \sum_{u_1=0}^{\infty} \sum_{u_2=0}^{\infty} q_{u_1} q_{u_2} \mathbb{P}(\mathcal{S}_i \leq u_1+1, \mathcal{S}_l \leq u_2+1) \\
&= \mathbb{E} \mathcal{M} + \mathbb{E}(\xi_1^2 - \xi_1) \sum_{i=1}^{\infty} \sum_{u=0}^{\infty} q'_u \mathbb{P}(\mathcal{S}_i \leq u+1) \\
&\quad + (\mathbb{E} \xi_1)^2 \sum_{i=1}^{\infty} \sum_{l=1}^{\infty} \sum_{u_1=0}^{\infty} \sum_{u_2=0}^{\infty} q_{u_1} q_{u_2} \mathbb{P}(\mathcal{S}_i \leq u_1+1, \mathcal{S}_l \leq u_2+1) \\
&\quad - (\mathbb{E} \xi_1)^2 \sum_{i=1}^{\infty} \sum_{u_1=0}^{\infty} \sum_{u_2=0}^{\infty} q_{u_1} q_{u_2} \mathbb{P}(\mathcal{S}_i \leq u_1 \wedge u_2+1). \tag{A.11}
\end{aligned}$$

Now, since  $\mathbb{E} \mathcal{N}_v = \sum_{i=1}^{\infty} \mathbb{P}(\mathcal{S}_i \leq v)$ ,  $\mathbb{E} \mathcal{N}_v = v / \mathbb{E} \tau_1$ ,

$$\mathbb{E} \mathcal{N}_{v_1} \mathcal{N}_{v_2} = \sum_{i=1}^{\infty} \sum_{l=1}^{\infty} \mathbb{P}(\mathcal{S}_i \leq v_1, \mathcal{S}_l \leq v_2),$$

we have

$$\sum_{i=1}^{\infty} \sum_{u=0}^{\infty} q'_u \mathbb{P}(\mathcal{S}_i \leq u+1) = \sum_{u=0}^{\infty} q'_u \mathbb{E} \mathcal{N}_{u+1} = \sum_{u=0}^{\infty} q'_u (u+1) / \mathbb{E} \tau_1 = [\mathbb{E}(\eta_{11} \wedge \eta_{12}) + 1] / \mathbb{E} \tau_1$$

and

$$\begin{aligned} & \sum_{i=1}^{\infty} \sum_{u_1=0}^{\infty} \sum_{u_2=0}^{\infty} q_{u_1} q_{u_2} \mathbb{P}(\mathcal{S}_i \leq u_1 \wedge u_2 + 1) \\ &= \sum_{u_1=0}^{\infty} \sum_{u_2=0}^{\infty} q_{u_1} q_{u_2} \mathbb{E} \mathcal{N}_{u_1 \wedge u_2 + 1} \\ &= \sum_{u_1=0}^{\infty} \sum_{u_2=0}^{\infty} q_{u_1} q_{u_2} \frac{u_1 \wedge u_2 + 1}{\mathbb{E} \tau_1} \\ &= [\mathbb{E}(\eta_{11} \wedge \eta_{12}) + 1] / \mathbb{E} \tau_1. \end{aligned}$$

Thus, it follows from (A.11) that

$$\begin{aligned} \mathbb{E}(\mathcal{M}^2) &= \mathbb{E} \mathcal{M} + [\text{Var}(\xi_1) - \mathbb{E} \xi_1] [\mathbb{E}(\eta_{11} \wedge \eta_{12}) + 1] / \mathbb{E} \tau_1 \\ &\quad + (\mathbb{E} \xi_1)^2 \sum_{u_1=0}^{\infty} \sum_{u_2=0}^{\infty} q_{u_1} q_{u_2} \mathbb{E} \mathcal{N}_{u_1+1} \mathcal{N}_{u_2+1}. \end{aligned} \quad (\text{A.12})$$

Define  $\Delta \mathcal{N}_i = \mathcal{N}_i - \mathcal{N}_{i-1}$ ,  $\Delta H(i) = H(i) - H(i-1)$ . For  $1 \leq m_1 < m_2$ , given  $\Delta \mathcal{N}_{m_1} = 1$ , we know a renewal has happened at time  $m_1$ , so

$$\mathbb{E}(\Delta \mathcal{N}_{m_1} \Delta \mathcal{N}_{m_2}) = \mathbb{E}(\Delta \mathcal{N}_{m_2} | \Delta \mathcal{N}_{m_1} = 1) \mathbb{P}(\Delta \mathcal{N}_{m_1} = 1) = \Delta H(m_2 - m_1) / \mathbb{E} \tau_1.$$

This yields

$$\begin{aligned} \mathbb{E}\{\mathcal{N}_u^2\} &= \mathbb{E} \left\{ \left( \sum_{m=1}^u \Delta \mathcal{N}_m \right)^2 \right\} \\ &= 2 \sum_{1 \leq m_1 < m_2 \leq u} \mathbb{E} \Delta \mathcal{N}_{m_1} \Delta \mathcal{N}_{m_2} + \sum_{m=1}^u \mathbb{E} \{(\Delta \mathcal{N}_m)^2\} \\ &= 2 \sum_{1 \leq m_1 < m_2 \leq u} \Delta H(m_2 - m_1) / \mathbb{E} \tau_1 + \sum_{m=1}^u \mathbb{E} \{ \Delta \mathcal{N}_m \} \\ &= (\mathbb{E} \tau_1)^{-1} \left( 2 \sum_{i=1}^{u-1} H(i) + u \right), \end{aligned} \quad (\text{A.13})$$

and, for  $u < v$ ,

$$\begin{aligned}
\mathbb{E}(\mathcal{N}_u \mathcal{N}_v) &= \mathbb{E}(\mathcal{N}_u^2) + \mathbb{E}[\mathcal{N}_u(\mathcal{N}_v - \mathcal{N}_u)] \\
&= \mathbb{E}(\mathcal{N}_u^2) + \sum_{m_1=1}^u \sum_{m_2=u+1}^v \mathbb{E}(\Delta \mathcal{N}_{m_1} \Delta \mathcal{N}_{m_2}) \\
&= \mathbb{E}(\mathcal{N}_u^2) + \sum_{m_1=1}^u \sum_{m_2=u+1}^v \Delta H(m_2 - m_1) / \mathbb{E} \tau_1 \\
&= (\mathbb{E} \tau_1)^{-1} \left( \sum_{i=1}^{u-1} H(i) + u + \sum_{i=v-u}^{v-1} H(i) \right). \tag{A.14}
\end{aligned}$$

Using (A.13) and (A.14), we have

$$\begin{aligned}
&\sum_{v_1=0}^{\infty} \sum_{v_2=0}^{\infty} q_{v_1} q_{v_2} \mathbb{E} \mathcal{N}_{v_1+1} \mathcal{N}_{v_2+1} \\
&= \sum_{v=0}^{\infty} q_v^2 \mathbb{E}(\mathcal{N}_{v+1}^2) + 2 \sum_{0 \leq v_1 < v_2} q_{v_1} q_{v_2} \mathbb{E} \mathcal{N}_{v_1+1} \mathcal{N}_{v_2+1} \\
&= (\mathbb{E} \tau_1)^{-1} \sum_{v=0}^{\infty} \left( 2 \sum_{i=1}^v H(i) + v + 1 \right) q_v^2 \\
&\quad + 2(\mathbb{E} \tau_1)^{-1} \sum_{0 \leq v_1 < v_2} \left( \sum_{i=1}^{v_1} H(i) + v_1 + 1 + \sum_{i=v_2-v_1}^{v_2} H(i) \right) q_{v_1} q_{v_2}.
\end{aligned}$$

This, together with (A.12), (A.1) and (A.2), gives (1). For (2), using the total probability formula, we have

$$H(n) = \sum_{k=1}^n \mathbb{E}(N_n | \tau_1 = k) \mathbb{P}(\tau_1 = k) = \sum_{k=1}^n (H(n-k)+1) f_k = \sum_{k=1}^n H(n-k) f_k + F(n),$$

as claimed.  $\square$

## Appendix B. Proof of Corollary 2

Let  $\mathcal{F}$  be the  $\sigma$ -algebra generated by  $\{\xi_i, i \geq 1\}$ , then we have

$$\begin{aligned}
\phi_{\mathcal{M}}(u) &= \mathbb{E} \left\{ \mathbb{E} \left( \prod_{i=1}^{\infty} \prod_{j=1}^{\xi_i} e^{iu \mathbf{1}_{\{\mathcal{S}_i \leq \eta_{ij} + 1\}}} \middle| \mathcal{F} \right) \right\} \\
&= \mathbb{E} \prod_{i=1}^{\infty} \prod_{j=1}^{\xi_i} (e^{iu} (1 - F_{\eta}(\mathcal{S}_i - 2)) + F_{\eta}(\mathcal{S}_i - 2)) \\
&= \mathbb{E} \prod_{i=1}^{\infty} (e^{iu} + (1 - e^{iu}) F_{\eta}(\mathcal{S}_i - 2))^{\xi_i} \\
&= \prod_{i=1}^{\infty} \psi_{\xi} (e^{iu} + (1 - e^{iu}) F_{\eta}(\mathcal{S}_i - 2)), \tag{B.1}
\end{aligned}$$

where  $\psi_{\xi}(s) = \mathbb{E} s^{\xi_1}$ . However,  $\xi_1 = \sum_{j=1}^{\zeta_1} I_{1j}$  is a mixed binomial with

$$\psi_{\xi}(s) = \psi_{\zeta}(r(s-1) + 1). \tag{B.2}$$

Hence, combining (B.1) and (B.2) gives (3). If  $\zeta_i \sim \text{Pn}(\lambda)$ , we have

$$\phi_{\mathcal{M}}(u) = e^{-r\lambda(1-e^{iu}) \sum_{i=1}^{\infty} (1-F_{\eta}(\mathcal{S}_i-2))},$$

which is the characteristic function of  $\text{Pn}(r\lambda \sum_{i=1}^{\infty} (1-F_{\eta}(\mathcal{S}_i-2)))$ .  $\square$

## Appendix C. Proof of Theorem 3

We have  $\mathbb{P}(\eta_{ij} \geq n + lD - i) = 0$  for  $n + lD - i > K$ , hence

$$M_{n+lD} = \sum_{i=1}^{n+lD} \sum_{j=1}^{\xi_i} \mathbf{1}_{\{\eta_{ij} \geq n+lD-i\}} = \sum_{i=n+lD-K}^{n+lD} \sum_{j=1}^{\xi_i} \mathbf{1}_{\{\eta_{ij} \geq n+lD-i\}} \text{ almost surely.}$$

The characteristic function of  $M_{n+lD}$  for  $n+lD > K$  is

$$\begin{aligned}
\phi_{n+lD}(t) &= \mathbb{E} \left\{ \mathbb{E} \left( \prod_{i=n+lD-K}^{n+lD} \prod_{j=1}^{\xi_i} e^{it \mathbf{1}_{\{\eta_{ij} \geq n+lD-i\}}} \middle| \xi_i, i = n+lD-K, \dots, n+lD \right) \right\} \\
&= \mathbb{E} \prod_{i=n+lD-K}^{n+lD} \prod_{j=1}^{\xi_i} (e^{it} (1 - F_{\eta,i}(n+lD-i-1)) + F_{\eta,i}(n+lD-i-1)) \\
&= \mathbb{E} \prod_{i=n+lD-K}^{n+lD} (e^{it} + (1 - e^{it}) F_{\eta,i}(n+lD-i-1))^{\xi_i} \\
&= \prod_{i=n+lD-K}^{n+lD} \psi_{\xi,i} (e^{it} + (1 - e^{it}) F_{\eta,i}(n+lD-i-1)) \\
&= \prod_{i=n+lD-K}^{n+lD} \psi_{\zeta,i} (r_i (e^{it} - 1) (1 - F_{\eta,i}(n+lD-i-1)) + 1),
\end{aligned}$$

where  $\psi_{\xi,i}(s) = \mathbb{E} s^{\xi_i}$  and the last equality is because the same reasoning as that for (B.2) gives

$$\psi_{\xi,i}(s) = \psi_{\zeta,i}(r_i(s-1) + 1).$$

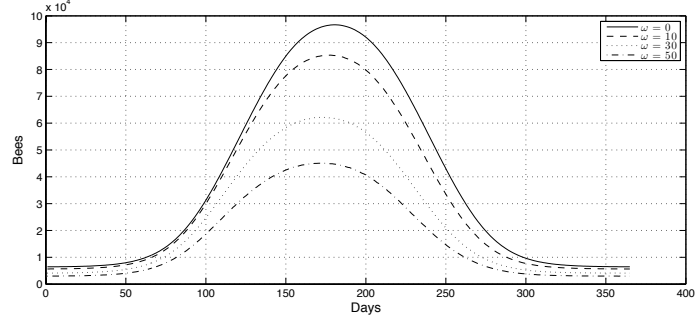
The mean of  $M_{n+lD}$  is a straightforward calculation. If  $\zeta_i \sim \text{Pn}(\lambda_i)$ , the characteristic function  $\phi_{n+lD}$  is reduced to

$$\phi_{n+lD}(t) = \exp \left( (e^{it} - 1) \sum_{i=n+lD-K}^{n+lD} r_i \lambda_i (1 - F_{\eta,i}(n+lD-i-1)) \right),$$

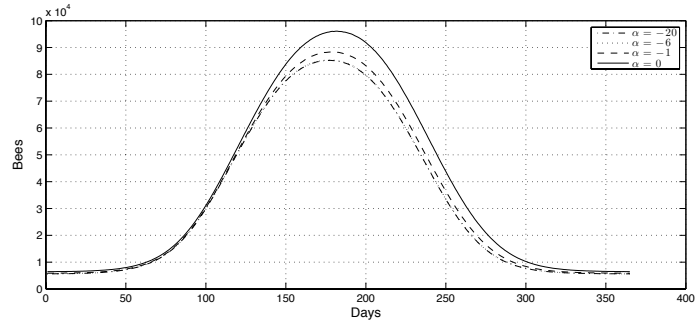
the same as the characteristic function of  $\text{Pn} \left( \sum_{i=n+lD-K}^{n+lD} r_i \lambda_i (1 - F_{\eta,i}(n+lD-i-1)) \right)$ .

□

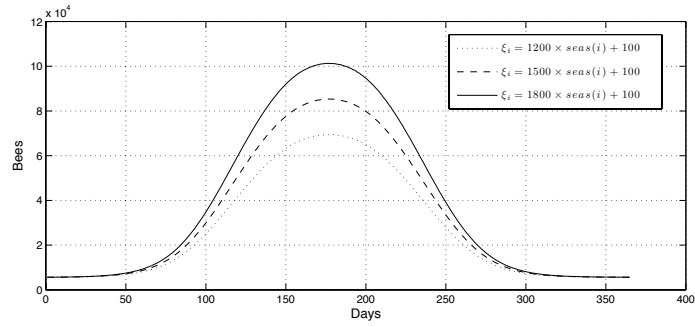
## Appendix D. Plots



a)



b)



c)

Figure D.6: a) The mean of the daily colony size  $M_n$  when  $\eta \sim SN(63, \omega^2, -6)$ , for the scale parameter  $\omega \in \{0, 10, 30, 50\}$ , b)  $\eta \sim SN(63, 10^2, \alpha)$ , for the slant parameter  $\alpha \in \{-20, -6, -1, 0\}$ , and a cyclic mean number of eggs:  $\mathbb{E}(\xi_i) = r_i \mathbb{E}(\zeta_i) = 1500 \times seas(i) + 100$ . c)  $\eta \sim SN(63, 10^2, -6)$ , and a cyclic mean number of eggs given by  $\mathbb{E}(\xi_i) = r_i \mathbb{E}(\zeta_i) = L \times seas(i) + 100$ ,  $L \in \{1200, 1500, 1800\}$ .